CHAPTER IX: THE BIOLOGICAL FUNCTION OF GLYCANS

Now, let me get to some of the key issues. I talked about the diversity of these glycans, I talked about where they're present. I talked about some of the fundamental, sort of the biochemical attributes, in terms of how they interact.

But, I want to put in front of you some of the key issues that get to the heart of structural functional relationship with these glycans, so that you can understand both the challenges and complexity and what the issues are to eventually be able to understand the biological function.

Perhaps there are two very important attributes which are listed here. First, unlike DNA and proteins, where you have a template for biosynthesis and assembly, carbohydrates do not have a template. They have a complex nontemplate biosynthesis. In other words, there is no open reading frame to give you a sense of what are the ways these sequences are assembled, and how do those sequences translate to function. And, equally important, you have proof reading processes for DNA. That doesn't exist for carbohydrates.

So, what happens as a consequence is, given the fact that there isn't a reading frame, carbohydrates are synthesized as an ensemble, the variety of different structures, and all these structures that get displayed on the cell surface are on the protein. So, that really begs the question, what are the relevant structures, or what is the distribution of the kind of structures, and, hence, what the functional consequences are.

Another important point is the fact that they cannot be amplified. What you isolate is what you have. And, typically, they're in low amounts. Therefore, it seriously poses some analytical challenges.

I mean, for instance, when you take molecules like heparin or chondroitin, once you isolate and you say that these are available in large amounts, there's still a fundamental issue of how you decouple this complexity of information in that mixture, and how does that mixture correlate to various functions?

Now, the point that I just elaborated on is the heterogeneous polydisburse high information content. Heparin has forty-eight possible different building blocks, and we do know that several of them are present in this mixture. And, part of the challenge that has really come in the way in understanding the structure function relationship is to demystify that and not look at them as individual species, but how do these ensemble of structures really mediate function? Therefore, you're trying to recapitulate the biology of these molecules as they are presented in the cell surface and extra cellular matrix so that, as against thinking about this in a simple one-structure one-sequence correlation to function, cells display a lot of these different molecules. They do parallel process a variety of signaling molecules lead to function. And, it sort of begs us to think about this in a very different way than we think about proteins and DNA.

And, perhaps the final point, which is very important, and, in some sense, also addresses the issue of why we have not been able to tackle this complex problem until now was hard, which is the fact that these molecules are present at the cell tissue interface. Therefore, we kind of need a more integrated way of understanding structure function relationship. And, what do I mean by that? I'm going to use the next slide to sort of really illustrate the key points here.

The fact that these molecules are present in the cell surface and the matrix and the fact that they do play a central role in how cells come together to form tissues in

organ systems, we needed to wait for several different technologies to come together. But, I wanted to put it in the context of historically biotechnology and molecular biology focused on variety of events that happened within the cells. How genes turned on and off, how various different signaling pathways regulated cell function. And, most emphasis was on the intercellular events and extracellular events and, particularly, given the fact that glycans are fairly the largest components of the extra cellular matrix, was viewed inaccessible, very likely to be inert material, structural material, much likely useless material. And, therefore, not important to study. But, we had to wait for whole organism genetics to really tell us something very interesting.

When people began studying flies and worms and looked at the process of morphogenesis, it became pretty clear that several of the aberrant phenotypes that they observed had to do with enzymes that played a direct or an indirect role with regard to regulating glycan biology. And, it then became pretty clear that a lot of the processes that truly happened happened in this interface. And, therefore, as against taking a more reductionist way, in order to truly understand the complexity of these molecules, we need to take a more integrated way, where we have to bring in this ensemble approach, so it's not a point to point comparison. How do you look at various structures interacting with various proteins, sort of a more mix approach, if you will, so that you can then begin to understand how various pathways are simultaneously affected, and it's sort of a very different way of thinking, in some sense, and, in part, it's also a way people are beginning to look at systems biology that, as against looking at individual pieces, how the components come together to define the complexity of biological processes. And, several different technologies are becoming available, high throughput approaches, more

integrated way of looking at data analysis and data integration and the power of informatics. And, these have truly become important to access glycol biology and understand the structure function relationships of glycol biology, and I'm going to walk through that in a more specific way to give you a sense of what that means.

So, just some quick thoughts on how to think about the tools and technologies to access structural information and structure-function activity.

Here's a schema again. Cell, cell surface, carbohydrates. Typically, you have chemical methods to generate a whole range of molecular weight distribution. Whether you're looking at heparin, chondroitin, there's an ensemble. The question is how do you take an integrated approach and deconvolute this to give you specific structures and sequences, their abundance and their location, so that you can begin to understand biology.

In many ways, I'm going to use this example to illustrate the concept behind how to approach sequencing these mixtures of complex carbohydrates. It's sort of like solving a puzzle during imaging. Here's a picture of an elephant. Use a variety of different techniques to image. You get a variety of different data sets for that. If you don't really integrate the data sets and look for correlation, your image is blurry. The more you have data, the more you're able to integrate, you're able to get a clearer picture of the image.

So, what do I mean by that? Part of the field in the area of carbohydrates that has really come up with this variety of orthoginal techniques, such as ESIMS, NMR, maldi MS, a variety of enzymes, to access not only the linear sugars but the branch sugars, so that you can then generate a variety of orthogonal data sets that can be brought

together to be able to look at various structures and how not only are you looking at sequences and components of sequences in a particular species, but look at this entire thing in the context of a mixture so that you can then begin to look at structure function relationship at a complex level, rather than a simple level, and we now have glycomic database that gets correlated with both genomic and proteomic database. This is a one sort of slide that really summarizes a variety of different things that different groups in the field have done, including ours, and we have specifically focused on the GAG field to be able to understand how to look at mixtures, looking at mixtures of heparin, mixtures of low molecular weight heparin, and mixtures of chondroitin as an example.

And, obviously, these have been very effecting in using glycans for biomarkers, because that's obviously a very important area. And, a lot of us do genetic and proteomic biomarkers, and the role of glycans, because you can access not only cells but serum to look at the various structures and the abundance, presence, absence, and ratios of the variety of these molecules, and this truly gets to some of the structure function relationship issues like I've been talking about.